International Journal of Information Science and Management
Vol. 14, No. 1, 2016, 47-56

The Evaluation of Individual Authors by Journal-Specific Metrics Would Not Yield Realistic Results

Hajar Sotudeh
Department of Knowledge & Information Sciences, Faculty of Education & Psychology
Shiraz University, Shiraz, Iran
Corresponding author: sotudeh@shirazu.ac.ir

Behnaz Soltani
Department of Knowledge & Information Sciences, Faculty of Education & Psychology
Shiraz University, Shiraz, Iran
Email: parsoltani@gmail.com

Abstract

To test the reasonability of applying journal-specific indicators with the purpose of evaluating individual researchers, the present study attempted to examine the structural similarities between journal-evaluation indicators (i.e. JIF, SNIP and SJR) and author-evaluation ones (i.e. publication counts, citation per paper, and H and G indices) through factor analysis. The Iranian papers having published in SCI in 2008 were chosen as the corpus of this study to be analyzed. The results showed that the author- and journal-evaluation indicators belong to two totally different factor groups, and share no structures. On this basis, one may conclude that what the journal- evaluation indices evaluate is completely different from what the author-level ones do. It would be, therefore, illogical to use these two groups of indices interchangeably and for purposes they have not been designed for. Otherwise, consistent results cannot be expected to come out of such endeavors.

Keywords: Scientific Production, Effectiveness, Author Evaluation, Journal Evaluation, Indicators.

Introduction

A variety of research evaluation indicators have been increasingly developed and applied to minimize the flaws of the previous measures and improve the validity and accuracy of evaluation results (Hirsch, 2005; Glänzel, 2006a; 2006b; Ortner, 2010; Moed, 2010). Unfortunately, however, the proliferation of the indicators leads to disorientation among decision-makers, who are no longer able to discriminate the pros and cons of the various indicators for planning an actual evaluation exercise (Abramo & D'Angelo, 2014). It is, therefore, necessary to refine the complicated evaluation mechanism and restore it to a more simple structure as expected from quantitative evaluation exercises.

Furthermore, the use of the indicators in contexts other than their original context may falsify the results of the evaluation. In particular, journal-evaluation indicators including IF, and the relatively recently-developed SJR and SNIP are used for the evaluation of individual researchers' scientific performance. This is in spite of the fact that each indicator has been devised to solve a particular problem, highlighting different characteristics or subtleties of a particular phenomenon (Pendlebury, 2009; Seglen, 1997) and despite the urgent need for research into new indices to evaluate their validity and advantages over their predecessors

(Butler, 2008; Harnad, 2008; Zitt & Bassecoulard, 2008). One may, therefore, call into question the reliability of such evaluations and the fairness of any decisions made on this basis.

According to previous studies, a general correlation has been found among author evaluation indicators, e.g. between the number of publications and citations (Lightfield, 1971; Cole & Cole, 1967; Katz, 1999), the H-index on the one hand and citation counts, publication counts, G-index and CPP, on the other (Saad, 2006; Van Raan, 2006; Cronin & Meho, 2006; Kelly & Jennions, 2006; Schreiber, 2010; Vanclay, 2008; Harzing & van der Wal, 2009; Tol, 2009). Journal-specific criteria are, also, found to be significantly correlated, e.g. impact factor and SJR, immediacy index and Eigen factor (Sadeghi & Sarraf Shirazi, 2012).

To the best of our knowledge, there exists no comprehensive study conducted at the author level trying to explore the underlying structure of journal- and author-specific indicators. As rare instances, one may notice the Leydesdorff's (2009) seminal work, and Costas and Bordons' (2007, 2008), which tried to examine the underlying structure of the scientometric indicators using factor analysis. However, the former has been conducted at the journal level and the two latter are limited to a narrow field i.e. Natural Resources. It is, therefore, necessary to study the interrelations of these two groups of indicators to not only refine the set of the indicators to form smaller journal- and author-evaluation scales, but also explore the applicability of journal- evaluation measures to the evaluation of individual researchers.

To do so, choosing Iranian researchers' scientific outputs indexed in SCI in 2008, the present study seeks to examine the structural similarities between journal-evaluation criteria (including Two-Year impact factor (2Y-IF), Five-Year Impact Factor (5Y-IF), SJR and SNIP) and individual-researcher-evaluation indicators (including Paper counts (P), Citation counts (C), Citation per Paper (CPP), H index and G index) via factor analysis and thereby to test the consistency of evaluation results based on the two groups of criteria.

Research methodology

Using citation analysis method, the present communication studies the Iranian corresponding authors indexed in SCI in 2008. The year 2008 was selected to ensure a five-year citation window (till OCT 2012, when the data were downloaded). The selection of the Iranian academic community is based on the fact that despite the call for cautious application of scientometric methods (Amani & Baba-Ahmadi, 2005; Davari Ardakani, 2007), it widely embraces such journal-evaluation indicators as IF or SNIP as a basis for decision-making about authors, with less than enough sensitivity or critical attitude. It is, thus, of utmost importance to adjust the country's research evaluation system, now just in its infancy.

Data Collection

A search was conducted in SCI in October 2012 using the search query CU=Iran limited to year 2008. This led to the identification of 15900 papers which were then downloaded in a tab-delimited format.

Identification of the Iranian Corresponding Authors: To identify the Iranian researchers, author-address field (i.e. RP) was parsed. First authors were identified by parsing C1 field, in cases where the former was left blank. The extracted data were then imported to Excel. After eliminating papers with non-Iranian corresponding authors and those carrying no author addresses, the number of papers to be studied reached down to 9711 papers.

To check the authors' variations and similarities in their names, their identities were verified in their online CVs and were also googled with their names, paper titles, organizational affiliations and email addresses. The standardization process led to the identification of 5515 unique researchers, with 168 having published in journals related to social sciences or journals with no subject category (See the section on "subject classification"). After excluding these researchers, the total number of researchers was reduced to 5347.

Identification of Researchers' Specialties: Classification of papers based on the related journals' subject categories is a common practice in scientometric studies. However, it could be challenging in studies carried out at the researcher level. Although expert in a particular field, a researcher may appear in several subject categories, depending on the classes of journals she published in. Given the fact that it is extremely difficult (if not impossible) to identify the researcher's main specialty using this method, a scientometrician would have to take all the related subject categories into account. By doing so, it would seem reasonable to distribute a single researcher's data among the subject categories of the related journals. For example, a researcher named Saeid Abbas-bandi with a total number of 12 papers and 392 citations, would have a CPP of 11 in chemistry, 5.50 in computer science, 69 in engineering, 30.80 in mathematics and 27 in physics. The scattering of a single researcher's data over more than one field in this way would not yield an accurate and comprehensive assessment of her whole performance. Furthermore, although this calculation is accurate and feasible in scientometric studies, it is too difficult and costly to implement and thus not recommendable for operational contexts of research evaluation where readymade and available data are needed for prompt decision-making.

An alternative solution, applicable both to scientometric studies and to the operational context of research evaluation, would be to take the researcher's whole data into account in every subject area she published in, without distributing the papers and citations among different subject areas. In the case of Saeid Abbas-bandi, for example, CPP would be set at 33, the H at 9 and the G at 12 for each and every subject area he appeared in. This method, too, suffers from its own defects. In other words, it would seem less accurate and might overestimate a researcher's performance in some fields, due to the inclusion of her whole data in the calculation for every subject area.

Using both of the classifications, the present study analyzed the data obtained from both methods. The analyses having yielded similar answers, only the results obtained from the first classification method is reported here for the sake of brevity.

Subject (Re)classification: There are over 249 subject categories (SC) defined in the Thompson Reuters databases, with 178 belonging SCI. The categorization being clearly too narrow, it may hinder discovering any possible trends across different subject areas. They were, therefore, reclassified into 19 ESI categories. It is worth mentioning that "Economics & Business"; "Psychiatry/Psychology" and "Social Sciences, general" being related to Social Sciences and Humanities (partly covered by SCI) were excluded. "Multidisciplinary

Sciences" were also eliminated from the study due to the small number of papers published by Iranian authors in the subject category.

Determining the indicators values: The H, G and CPP values were manually calculated for each researcher. The impact values of 2437 journals constituting Iran's publication strategy were extracted from JCR and Scopus. A very small percentage of Iranian papers were found to have been published in 101 journals (4.14%) with no IFs (either 2-Y or 5-Y IF) and in 346 titles (10.01%) with no SNIP or SJR, which were excluded from the study.

Data Analysis Method

The underlying structures of the author-specific and journal-specific evaluation indices were investigated using factor analysis. Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy showed adequate fit (KMO=0.753), exceeding the recommended value of 0.6. The Bartlett's test of sphericity was high at 4.145 and reached statistical significance (p=0.001), supporting the factorability of the correlation matrix. The Oblimin Rotation method with Kaiser normalization was used to extract the factors.

Findings

Structural Similarity between researcher- and journal-evaluation criteria

The results of factor analysis for determining the factor loadings of the evaluation criteria revealed that the variables clearly belong to two distinct components (T1). The first component includes P (r=0.891), C (r=0.918), H (r=0.948) and G (r=0.970), all being devised to evaluate individual researchers. The second component includes 2-Y IF (r=0.937), 5Y-IF (r=0.945), SJR (r=0.879) and SNIP (r=0.536), all being among journal-specific evaluation indicators.

As seen in Table 1, CPP is the one and only indicator that indicates significant, though relatively slight, factor loadings on both components. The G-index with a significant factor loading of 0.97, and the 2Y-IF and 5Y-IF index with factor loadings of 0.937 and 0.945, were found to be the strongest among the researcher- and journal-evaluation indices respectively.

Table1

The factor loadings of the evaluation indicators

Indicator	Component	
	1	2
P	0.891	
C	0.918	
CPP	0.466	0.396
H-Index	0.948	
G-Index	0.970	
2Y-IF		0.937
5Y-IF		0.945
SJR		0.879
SNIP		0.536

Structural Similarity between Journal- and Author-evaluation Indices in Various Disciplines

Table 2 shows the results obtained from the factor analysis of the indicators for various disciplines. As shown in Table 2, the P, H and G indicators show strong loadings on the first group of indices, that is, the author-evaluation indices. The same is true about the 2Y-IF, 5Y-IF, SJR and SNIP exhibiting significant and strong loadings on the second group, i.e. journal-evaluation indices. The total number of citations (C) shows factor loadings on both groups in Agricultural Science (r=0.788, r=0.569), Microbiology (r=0.537, r=0.678) and Plant & Animal Science (r=0.570, r=0.744).

The CPP has the same status in both groups in Engineering (r=0.532, r=0.541) and Geosciences (r=0.572, r=0.509). The indicator, however, has no significant factor loading on any of the components in Biology, Biochemistry, Mathematics, Molecular Biology, Genetics and Space Science. Furthermore, it shows factor loadings on journal-evaluation indices in some subject areas as Agricultural Science, Molecular Biology & Genetics and Immunology, while having factor loadings on researcher-evaluation indices in some others e.g. Pharmacology, Psychiatry/Psychology and Material Science.

As indicated by the table, the G-index shows the strongest factor loading on the first component, i.e. researcher-evaluation indices, in all subject areas; while the 2Y-IF or 5Y-IF show the strongest factor loadings on the second one- i.e. the journal-evaluation indices- in most of the subject areas. Meanwhile, the factor loadings associated with the SJR and SNIP indices are very high in most of the subject areas, exceeding 0.80.

Table2

The indicators factor loadings in different disciplines

Field	Indicator	Component		Field	Indicator	Component		Field	Indicator	Component	
		1	2			1	2			1	2
Agricultural Sciences	P	0.803		Clinical Medicine	P	0.665		Environment/Ecology	P	0.912	
	C	0.788	0.569		C	0.895			C	0.869	
	CPP		0.733		CPP	0.656			CPP		0.608
	H	0.928			H	0.919			H	0.953	
	G	0.95			G	0.935			G	0.966	
	2Y-IF		0.967		2Y-IF		0.971		2Y-IF		0.947
	5Y-IF		0.973		5Y-IF		0.977		5Y-IF		0.969
	SJR		0.962		SJR		0.946		SJR		0.914
	SNIP		0.944		SNIP		0.961		SNIP		0.817
Biology & Biochemistry	P	0.814		Computer Sciences	P	0.829		Geosciences	P	0.571	
	C	0.769			C	0.894			C	0.752	
	CPP				CPP	0.516			CPP	0.572	0.509
	H	0.916			H	0.894			H	0.913	
	G	0.942			G	0.956			G	0.926	
	2Y-IF		0.952		2Y-IF		0.903		2Y-IF		0.87
	5Y-IF		0.949		5Y-IF		0.961		5Y-IF		0.826

Field	Indicator	Component		Field	Indicator	Component		Field	Indicator	Component	
		1	2			1	2			1	2
	SJR		0.898		SJR		0.936		SJR		0.875
	SNIP		0.769		SNIP		0.915		SNIP		0.767
Chemistry	P	0.921		Engineering	P	0.792		Immunology	P	0.785	
	C	0.941			C	0.831			C	0.946	
	CPP		0.614		CPP	0.531	0.541		CPP	0.706	
	H	0.955			H	0.934			H	0.924	
	G	0.978			G	0.948			G	0.961	
	2Y-IF		0.964		2Y-IF		0.914		2Y-IF		0.971
	5Y-IF		0.974		5Y-IF		0.912		5Y-IF		0.985
	SJR		0.965		SJR		0.897		SJR		0.949
	SNIP		0.899		SNIP		0.863		SNIP		0.92

Table 2
The indicators factor loadings in different disciplines (continued)

Field	Indicator	Factor		Field	Indicator	Factor		Field	Indicator	Factor	
		1	2			1	2			1	2
Materials Science	P	0.872		Molecular Biology & Genetics	P	0.749		Physics	P	0.846	
	C	0.904			C	0.86			C	0.892	
	CPP	0.533			CPP				CPP	0.527	
	H	0.94			H	0.922			H	0.922	
	G	0.956			G	0.931			G	0.961	
	2Y-IF		0.897		2Y-IF		0.978		2Y-IF		0.939
	5Y-IF		0.946		5Y-IF		0.992		5Y-IF		0.956
	SJR		0.921		SJR		0.963		SJR		0.932
	SNIP		0.8		SNIP		0.943		SNIP		0.815
Mathematics	P	0.87		Neurosciences & Behavior	P	0.75		Plant & Animal Science	P	0.701	
	C	0.921			C	0.913			C	0.57	0.744
	CPP				CPP	0.519			CPP		0.66
	H	0.948			H	0.943			H	0.923	
	G	0.97			G	0.954			G	0.944	
	2Y-IF		0.908		2Y-IF		0.965		2Y-IF		0.972
	5Y-IF		0.92		5Y-IF		0.942		5Y-IF		0.946
	SJR		0.842		SJR		0.968		SJR		0.958
	SNIP		0.873		SNIP		0.814		SNIP		0.896
Microbiology	P	0.81		Pharmacology	P	0.897		Psychiatry/Psychology	P	0.937	
	C	0.537	0.678		C	0.893			C	0.949	
	CPP		0.695		CPP	0.554			CPP	0.515	
	H	0.932			H	0.938			H	0.92	
	G	0.956			G	0.948			G	0.955	

Field	Indicator	Factor		Field	Indicator	Factor		Field	Indicator	Factor	
		1	2			1	2			1	2
	2Y-IF		0.955		2Y-IF		0.959		2Y-IF		0.953
	5Y-IF		0.957		5Y-IF		0.972		5Y-IF		0.956
	SJR		0.955		SJR		0.957		SJR		0.959
	SNIP		0.895		SNIP		0.888		SNIP		0.83
Space Sciences	P	0.876									
	C	0.836									
	CPP										
	H	0.905									
	G	0.948									
	2Y-IF		0.892								
	5Y-IF		0.891								
	SJR		0.909								
	SNIP		0.852								

Discussion and Results

The results obtained from factor analysis indicate that the total number of publications, the total number of citations, H and G indices (with factor loadings of 0.89, 0.91, 0.94 and 0.97, respectively) have a common structure and can be grouped together as “author-evaluation indicators”. Furthermore, 2Y-IF, 5Y-IF, SJR and SNIP (with factor loadings of 0.93, 0.94, 0.87 and 0.53, respectively), all being essentially devised for “journal evaluation”, share common structures as well.

However, the so-called “author-evaluation indicators” and “journal-evaluation indicators” have no structures in common (Table 2). In line with the findings, Costas and Bordons (2007; 2008) found that the total number of publications, the total number of citations, H and G indices belong to the same group of indices. Furthermore, in a study by Leydesdorff (2009), the total number of publications and citations were classified into one group, while the IF and SJR were classified as belonging to another group.

Given the values of factor loadings obtained in the present study, the G and H indices are respectively the strongest among the researcher-evaluation indices although the number of publications and the number of citations also indicate a significant factor loading. Furthermore, the 2Y-IF, 5Y-IF and SJR are respectively strong among all the journal-evaluation indices.

Among the researcher-evaluation indices, only the CPP shows factor loadings on both components. The lack of a strong correlation between CPP and other indicators shows that it yields considerably different results from the other ones even though all are founded on publications and citations.

Factor analysis in subject areas yielded similar results, though with some exceptions. For instance, the total number of citations in Agricultural Science, Microbiology and Plants & Animal Science shows significant factor loadings on both the author- and journal-evaluation indicators. Furthermore, in some subject areas like Engineering and Geology, CPP shows factor loadings on both components, while showing no factor loading on none of them in some other fields such as Biology & Biochemistry, Mathematics, Molecular Biology, Genetics and Space Science. In addition, in some subject areas like Agricultural Science, Ecology/Environment, Animal & Plant Science, Chemistry and Microbiology, the CPP shows a strong factor loading on the journal-evaluation indices. This finding provides further evidence on citation behavior differences among subject areas, and the necessity of vigilance when comparing researchers in different disciplines.

Conclusion

Accurate and impartial evaluation of researchers can play a crucial role in optimizing the decision and policy-making processes in academic contexts. It should now have become clear that there is no structural similarity between the so-called journal-evaluation and author-evaluation indices. This further highlights the fact that each index, designed for a particular purpose, cannot necessarily evaluate all scientific entities. The two groups of indices evaluate two completely distinct phenomena and consequently it would be misleading to use them interchangeably.

Using numerous indicators in evaluating researchers' performances is not only costly, but also complicated and time-consuming. Therefore, it would be desirable to have a smaller and more simple set of indices. The results of the present study showed that the G-index is the strongest of all the researcher-evaluation indices whether at the subject level or in general. It can, thus, obviate the need for using multiple indices. Among the journal-evaluation indices the 2Y-IF or 5Y-IF indices showed the strongest factor loadings in most subject areas. However, since the SNIP and SJR have shown strong factor loadings in most cases, we apparently can apply each of the indices to evaluate journals and indulge a hope of getting fair outcomes.

References

Abramo, G., & D'Angelo, C. A. (2014). Research evaluation: improvisation or science? Retrieved Sep 12, 2014 from http://www.portlandpress.com/pp/books/online/wg87/087/0055/0870055.pdf.

Amani, M. & Baba-Ahmadi, A. (2005). The inefficiency of JIF in evaluating papers and research outcpmes. Rahyaft [Persian], 15 (36): 70-76.

Butler, L. (2008). Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research quality Framework. Ethics in Science and Environmental Politics ESEP, (8).

Cole, S. & Cole, J. R. (1967). Scientific output and Recognition: Astudy in the operation of the reward system in Science. American Sociological Review, 32(3), 377-390.

Costas, R. & Bordons, M. (2007). The h-index: Advantages, limitations and its relation withother bibliometric indicators at the micro level Journal of Informetrics, (1), 193–203.

Costas, R .& Bordons, M. (2008). Is g-index better than h-index? An exploratory study at the individual level. Scientometrics, (77)2, 267–288.

Cronin, B. & Meho, L. (2006). Using the h-index torank influential information scientists. Journal of the American Society for InformationScience and Technology, 57(9), 1275-1278.

Davari Ardakani, R. (2007). The delusion of scientific development via increasing ISI papers. Iran Newspaper [Persian], 3710, 10.

Glanzel, W. (2006a). On the opportunities and limitations of the H-index. Science Focus, *Library of Chinese Academy of Sciences*, 10-11. (Published)[Newspaper/Magazine article].

Glanzel, W. (2006b). On the h-index - A mathematical approach to a new measure of publication activity and citation impact. *Scientometrics*, 67 (2), 315-321.

Harnad, S. (2008). Validating Research Performance Metrics Against Peer Rankings, *Ethics in Science and Environmental Politics*, 8 (11).

Harzing, A. W. & Van der Wal, R. (2009). A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. *Journal of the American Society for Information Science and Technology*, 60(1), 41-46.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research Output. Proceedings of the National Academy of Sciences of the UnitedStates of America, 102(46), 16569–16572.

Katz, J. S. (1999). Bibliometric Indicators and the Social Sciences, prepared for the ESRC. Available online at: http://arizona.openrepository.com/arizona/bitstream/10150/105920/1/ESRC.pdf

Kelly, C. D. & Jennions, M. D. (2006). The h index and career assessment by numbers. *Trends in Ecology and Evolution*, 21(4):167-170.

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? *Journal of the American Society for Information Science & Technology*, 60, 1327–1336.

Lightfield, E. T. (1971). Output and Recognition of Sociologists. *The American Sociologist*, 6(2), 128-133.

Moed, H. F. (2010). The Source-Normalized Impact per Paper (SNIP) is a valid and sophisticated indicator of journal citation impact. Digital libraries, cite as: arXiv: 1005. 4906v1

Ortner, H. M. (2010). The impact factor and other performance measures-much used with little knowledge about. *International Journal of Refractory Metals and Hard Materials*, 28(5), 559-566.

Pendlebury, D. A. (2009). The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1), 1-11.

Saad, G. (2006). Exploring the h-index at the author and journal levels using bibliometric data of productive consumer scholars and business-related journals respectively. *Scientometrics*, 69(1), 117-120.

Sadeghi, R. & Sarraf Shirazi, A. (2012). Comparison between Impact factor, SCImago journal rank indicator and Eigenfactor score of nuclear medicine journals. *Nuclear Medicine Review*, 15(2), 132–136

Schreiber, M. (2010). Revisiting the g-Index: The average number of citations in the g-core. *Journal of the American Society for Information Science and Technology*, 61(1),169-174.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, (314),498-502. Retrieved 12 Sep 2013 from http://bmj. bmjjournals. com/cgi/content/full/314/7079/497.

Tol, R. S. J. (2009). The h-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics*, 80(2), 317-324.

Vanclay, J. K. (2008). Ranking forestry journals using the h-index. *Journal of informetrics*, 2(4), 326-334.

Van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491-502.

Zitt, M.; Bassecoulard, E. (2008). Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. Ethics in Science and environmental Politics Ethics Sci Environ Polit, (8), 49–60.